

Notat om implementering af De Nationale Test

Nye initiativer, som staten beslutter (fx folkeskolereformen i 2013, nye Fælles Mål i 2009 eller De Nationale Test i 2010), skal føres ud i livet af nogle andre (fx kommuner eller skoler), end dem der beslutter initiativet. Tre grundlæggende elementer spiller en rolle for succes:

- 1) *Er ideen god:* Er initiativ/vejledninger/intentioner gennemtænkt og rigtigt formulerede?
- 2) *Virker det.* Fungerer de tekniske/administrative løsninger som forudsat?
- 3) *Implementeringen.* Er initiativet ført ud i livet med de redskaber/signaler, som der var tænkt og beskrevet?

Opmærksomheden henledes nogle gang på de to første, men hvor det ofte er den tredje – implementeringen – der har den største betydning for virkningen, og den kan opgøres med de rigtige redskaber.

Implementeringsgraden, eller implementeringsfideliteten, for De Nationale Test er emnet for dette notat.

Baggrund

Efter vedtagelsen af de nationale test (DNT) i 2006 og indførelsen af dem i fuld skala (i 2010) har implementeringen foregået meget forskelligt fra kommune til kommune og fra skole til skole, både hvad angår hvem der bestemmer hvad, hvordan testene afvikles, hvordan resultaterne bliver brugt og til hvilke formål. Ved vedtagelse af loven var der en række intentioner med at gennemføre DNT sammen med signaler om, hvordan de skulle føres ud i livet. En del af disse overvejelser kan læses af lovforslagets omfattende og detaljerede bemærkninger, jf. følgende uddrag:

”Med henblik på at følge den enkelte elevs tilegnelse af kundskaber og færdigheder, herunder styrker, svagheder og potentialer, foreslås det, at der som led i evalueringen af elevens udbytte af undervisningen skal anvendes centralt udarbejdede test i udvalgte fag på bestemte klassetrin.

Testene er et pædagogisk værktøj. De skal anvendes sammen med resultaterne af evalueringen i øvrigt fremadrettet til brug for den videre planlægning af undervisningen, i vejledningen af den enkelte elev, og i underretningen af forældrene med henblik på at tilrettelægge en undervisning og et forældresamarbejde, der understøtter eleven bedst muligt.” (Kilde: Lovforslag L101, samling 2005/06¹)

I den offentlige debat diskuteres testene som regel uden, at der ses nærmere på, hvordan de reelt anvendes, og hvad testresultaterne bruges til. Et spørgsmål, som trænger til afklaring, er derfor, om skolerne og kommunerne har implementeret testene hensigtsmæssigt. Det vil der være mange synspunkter på, hvad der vil sige, men for at få en fast reference, ser vi på praksis i forhold til, hvordan testene oprindeligt var tænkt jf. lovens bemærkninger samt notat fra UvM².

Hvad er implementeringsfidelitet?

¹ Se http://www.ft.dk/samling/20051/lovforslag/l101/20051_l101_som_fremsat.htm

² [Notat fra UvM om de nationale test som pædagogisk værktøj \(12.09.2016\)](#)

Der er udviklet metoder til at vurdere implementeringsgraden på en skala ved at analysere implementeringen med afsæt i hensigten med interventionen. Denne implementeringsgrad kaldes *implementeringsfidelitet*³.

En implementeringsfidelitet er altså et mål for, i hvilken grad man ved implementering af en intervention har været trofast mod de oprindelige tanker om midler og processer ved implementering. Den korte udgave lyder: ***I hvilket omfang er en intervention leveret, som den er tænkt?*** En høj implementeringsfidelitet er således kun en garanti for et vellykket projekt, hvis projektet er rigtigt tænkt fra start til slut. Man kan således sagtens forestille sig vellykkede projekter med lav implementeringsfidelitet og omvendt fejlslagne initiativer med høj implementeringsfidelitet.

Metoden med at score implementering er oprindeligt udviklet til forskningsbrug, hvor man typisk forsøger at undersøge interventioners effektivitet ved brug af RCT⁴-inspirerede designs. Men redskaber af denne type er også godt egnede til at følge med i, hvordan det står til i en lokal implementeringsproces, hvor man kan have brug for hjælp til at følge sin egen udvikling for at se, hvor langt man er fra at leve op til den implementeringsstrategi og de mål, der var sigtet for den oprindelige hensigt. Især når det drejer sig om komplekse problemstillinger med komplekse mål.

Hvad siger staten?

Ved indførelsen af de nationale test er der dels i lovforslagets bemærkninger og dels i det vejledningsmateriale, der skulle støtte op om implementeringen, afgivet markante signaler om, hvad der er god praksis. De råd og anvisninger, der er givet, skal kort gennemgås:

Det blev fra starten meldt ud, at lærerne bør *instruere eleverne og afvikle testene* på en sådan måde, at eleverne ikke oplever nederlag, og så man får retvisende data. F.eks. blev det understreget, at eleverne skal kende vilkårene (fx hvilke hjælpemidler, der anvendes, testopgavernes formater, at opgaverne vil blive udvalgt til elevens faglige niveau, og at testen fortsætter, indtil læreren stopper den), og at læreren skulle gøre sit bedste for at sikre ro og fokus på det faglige (fx undgå at vise lærerens oversigtsskærm på whiteboard under testen, da det medfører konkurrence om at svare flest opgaver).

Fortrolighed om testresultater blev besluttet for at undgå, at testene blev et styringsredskab (fx til brug ved resultatløb, offentlig rangordning, summativ anvendelse), men det har betydet, at man flere steder har undladt at give lærerne systematisk adgang til relevante testresultater – dvs. både de seneste og de tidligere resultater for deres elever. Det har været påpeget, at det er ledelsens opgave at sikre, at *alle lærere har adgang til og dermed mulighed for* at analysere deres elevers besvarelser. Dette (at læreren kan bruge testene til at følge den enkelte elevs tilegnelse af kundskaber og færdigheder) er en helt grundlæggende forudsætning for, at testene kan anvendes formativt/til pædagogiske formål.

Både de obligatoriske og frivillige test bør tilrettelægges, så resultaterne kan bruges fremadrettet til brug for den videre planlægning af undervisningen – altså formativt. Det betyder, at der bør testes *før* undervisningen er blevet gennemført – fx tidligst muligt i den obligatoriske periode. Dette krav er ud fra internationale erfaringer vanskeligt at formidle til skoler, der i højere grad er vant til, at test og prøver bruges som et summativt redskab til brug for måling af skolens effektivitet – dvs. at der testes senest muligt – *efter* undervisningen er blevet gennemført. DNT blev indrettet således, at eleverne i dansk/læsning, matematik og nu også engelsk kan testes én gang årligt i relevante klassetrin med henblik

³ Begrebet fidelitet kommer fra fransk *fidèle* 'trofast, loyal' og fra latin *fidelis* 'trofast'.

⁴ Kontrolleret eksperiment (RCT=Randomized controlled trial).

på at følge elevens faglige udvikling over lange stræk. Dette fordrer en intelligent anvendelse af de frivillige test. Mange steder har skolerne valgt at give frit, hvad testene anvendes til – og ofte er de frivillige test blevet brugt som øvelse, hvilket altså er i indirekte konflikt med formålet for testene. Eftersom læreren ikke (grundet fortrolighed om opgaver) tillades at give feedback til elever på konkrete opgaver, er det en nytteløs måde at øve på.

Et væsentligt formål med testene er feedback, ikke blot til eleven, men også til forældre. Det skal ske på en måde, der understøtter elevens udvikling bedst muligt. Det betyder, at resultaterne skal formidles sammen med planer om, hvordan potentialer understøttes, og udfordringer adresseres. Eleven bør have resultater af test direkte efter testens afvikling (fx inden for en uge) – således at vedkommende får en melding på sin præstation, sin udvikling og de videre planer, mens testen stadig er i frisk erindring. Forældrene bør have feedback om samme emner på førstkommende skole-hjem-samtale (gerne hurtigt efter), og der bør følge råd med om, hvordan forældrene bedst understøtter deres barns udvikling, idet der kobles til anden evaluering. Det bør ikke fylde ret meget i skole-hjem-samtalen (højst et par minutter – hvis der ikke er særlige problemer eller forhold).

Mange skoler har ladet det være op til de enkelte lærere at bestemme anvendelsen af DNT. Det betyder, at data er indsamlet uegalt, og resultater har været umulige at fortolke præcist. Fra centralt hold er der gjort en vis indsats for at afhjælpe situationen⁵, men der er behov for en koordineret indsats og planlægning, uddannelse og support (på skole og/eller kommuneniveau), for at kunne anvende testene som forudsat i lovforslaget. I flere kommuner/skoler bliver der udarbejdet centrale retningslinjer for pædagogisk anvendelse af testene, og det er stadig mere udbredt, at der ydes en betydelig indsats for at især vejlederne (fx i læsning / matematik / læring / it) kan supportere ledelse og lærere vedr. DNT.

Det stod helt centralt i lovforslaget, at skolen skulle benytte resultaterne systematisk i en stærk lokal evalueringskultur. Det har vist sig meget mere vanskeligt at implementere end først antaget. Den oprindelige ekspertanalyse fra OECD-teamet i 2004⁶ pegede på, at danske skoler har en meget svag kultur for evaluering, og det var en del af missionen med DNT at ændre denne kultur. Imidlertid kræver det en stærk evalueringskultur at udnytte et så avanceret redskab, som DNT udgør – det har muligvis medvirket til en relativ langsom implementering og en begrænset implementeringsgrad (jf. fx Rambøll 2013).

En grundlæggende forudsætning for, at DNT kan anvendes formativt, er, at der er klarhed om, at testene bruges formativt til undersøgelse af elevernes faglige niveau – ikke summativt til at bedømme læreres og skolars indsats med det formål at skaffe data til brug for belønning, formaninger eller straf (accountability). Internationale erfaringer viser, at lærere og skoler kan påvirke testens resultater på en sådan måde, at det vil være ødelæggende for testens validitet og have en meget kedelig indflydelse på undervisningen⁷. Skiftende regeringers anvendelse af testresultater som grundlag for styring

⁵ fx har der - til lærerens information - været en markering på lærerens monitorskærm af hvornår testen har passeret grænsen for den maksimalt acceptable usikkerhed: SEM=0,55. Når SEM på estimatet under testen kom ned under 0,55 blev markøren grøn" udfør elevens navn. Dette fortolkede mange lærere fejlagtigt som at nu bør testen stoppes – uanset at den kun havde kørt ganske kort tid. Denne misforståelse lykkedes det ikke at vejlede sig ud af, så ministeriet valgte en teknisk løsning: Først at lade testen "gå i grøn" når der er gået 45 minutter.

⁶ Se: <http://static.uvm.dk/publikationer/2004/oece/hel.html>

⁷ Campbells lov siger at: "Achievement tests may well be valuable indicators of general school achievement under conditions of normal teaching aimed at general competence. But when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways". Campbell, D. T. (1979). Assessing the impact of planned social change. Evaluation and program planning, 2(1), 67-90.

(Folkeskolereformens succeskriterier) har skabt uklarhed om den pædagogiske målsætning. Imidlertid er der en stærk vilje i skolerne og kommuner til ikke at lade sådanne politiske meldinger føre sig ud af kurs.

Ikke desto mindre forudsætter det en god balance og fælles målsætning i kommunen at holde rygter og spøgelse på afstand. Stærkt fokus på kvalitetsrapporter og gennemsnitsresultater kan fjerne fokus fra den fælles interesse, der er i at gøre eleverne så dygtige, som de kan – også målt efter andre standarder end test og prøver. Det er essentielt, at kommunalbestyrelsen, forvaltning, skoleledelse og lærere har gensidig forståelse og tillid til, at der er fælles interesse i en pædagogisk anvendelse af testene – denne tillid er ikke altid til stede, uanset at der som regel er identiske målsætninger på alle niveauer.

Implementeringen af DNT har været langsommelig og præget af usikkerhed – der er i processen kommuner og skoler, som har mistet tilliden til, at det nytter. De fleste steder er det gået fremad, men ikke alle. Enkelte kommuner har ligefrem bedt om dispensation fra at skulle anvende testene obligatorisk. Det kan der være mange årsager til: En af de væsentligste er nok, at der i debatten om testene er sejlivede udokumenterede rygter om testenes usikkerhed og anvendelighed. Men de fleste steder er der klart formulerede ønsker om og vilje til at udvikle på anvendelsen af DNT hen imod det formative, pædagogiske redskab, der var intentionen i loven fra 2006.